Available online at www.sciencedirect.com

SCIENCE DIRECT®

Bioorganic & Medicinal Chemistry 13 (2005) 1119–1129

Bioorganic & Medicinal Chemistry

Predicting multiple drugs side effects with a general drug-target interaction thermodynamic Markov model

Humberto González-Díaz,[a,b] Maykel Cruz-Monteagudo,[b,c,*] Reinaldo Molina,[b,d] Esvieta Tenorio[b] and Eugenio Uriarte[a]

[a]*Department of Organic Chemistry, Faculty of Pharmacy, University of Santiago de Compostela 15782, Spain*
[b]*Chemical Bioactives Centre, Central University of 'Las Villas' 54830, Cuba*
[c]*Applied Chemistry Research Centre, Central University of 'Las Villas' 54830, Cuba*
[d]*Universität Rostock, FB Chemie, Albert-Einstein-Street 3a, D 18059 Rostock, Germany*

Received 4 November 2004; revised 9 November 2004; accepted 12 November 2004
Available online 28 December 2004

Devoted to Markov, A. A. due to his outstanding work: Markov, A. A. *Bull. Soc. Phys. Math. Kasan.* **1906**, *15*, 155

Abstract—Most of present molecular descriptors just consider the molecular structure. In the present article we pretend extending the use of Markov chain models to define novel molecular descriptors, which consider in addition to molecular structure other parameters like target site or toxic effect. Specifically, this molecular descriptor takes into consideration not only the molecular structure but the specific system the drug affects too. Herein, it is developed a general Markov model that describes 39 different drugs side effects grouped in 11 affected systems for 301 drugs, being 686 cases finally. The data was processed by linear discriminant analysis (LDA) classifying drugs according to their specific side effects, forward stepwise was fixed as strategy for variables selection. The average percentage of good classification and number of compounds used in the training/predicting sets were 100/100% for systemic phenomena (47 out of 47)/(12 out of 12) and metabolic (18 out of 18)/(5 out of 5), muscular–skeletal (23 out of 23)/(6 out of 6) and neurological manifestations (33 out of 33)/(8 out of 8); 97.6/96.7% for cardiovascular manifestation (122 out of 125)/(30 out of 31); 97.1/97.5% for breathing manifestations (34 out of 35)/(8 out of 9); 97/99.4% for gastrointestinal manifestations (159 out of 164)/(40 out of 41); 97/95% for endocrine manifestations (32 out of 33)/(7 out of 8); 96.4/94.6% for psychiatric manifestations (53 out of 55)/(13 out of 14); 95.1/99.1% for hematological manifestations (98 out of 103)/(25 out of 26) and 88/92.3% for dermal manifestations (44 out of 50)/(12 out of 13). In addition, we report preliminary experimental reversible decrease of lymphocytes differential count after administration of the antibacterial drug G-1 in mice, which coincide with a posterior probability ($P\% = 74.91$) predicted by the model. This article develops a model that encompasses a large number of side effects grouped in specific organ systems in a single stochastic framework for the first time.

1. Introduction

At present there are more than 15 million chemical compounds that have been discovered or synthesized in chemical laboratories. A great quantity of these compounds have not found pharmacological or agrochemical applications yet. This is a consequence of the difference that exists between the rate at which novel chemical are discovered each year and the number of compounds that can be tested in chemical or pharmacological assays. These kinds of assays, specially pharmacological and toxicological ones, are in general expensive and time consuming. However, novel paradigms for drug discovery have been introduced recently, based on the ability of large chemical libraries and robotic system for bioassays. This system of high-throughput biochemical assays allow for the synthesis and testing of hundreds of compounds each day.[1]

During last year, the pharmaceutical industries have reoriented their research strategies in order to give more attention to those methods that permit the 'rational' selection or design of novel compounds with the desired properties.[2–4] In this sense, quantitative structure–toxicity relationships (QSTRs) are used as predictive tools for a preliminary evaluation of the hazard of chemical

Keywords: Drugs side effects; Markov model; Thermodynamics.

* Corresponding author. Tel.: +53 42 281473; fax: +53 42 281455; e-mail addresses: maikelcm@uclv.edu.cu; maikelx77@yahoo.es

0968-0896/$ - see front matter © 2004 Elsevier Ltd. All rights reserved.
doi:10.1016/j.bmc.2004.11.030

compounds by using computer aided models.[5–10] These theoretical models represent an alternative to the 'real' world of assaying chemical compounds for determining their toxicological properties on live organisms in the laboratory avoiding the expensive, time-consuming and in many cases animal aggressive bioassays, which are now made only after preliminary predictions with computational models.[11–16]

In general, González and co-workers, have recently discussed that QSTR studies can be applied to congeneric and noncongeneric data sets of compounds. The first permits the understanding of specific toxicological mechanism of action for molecules structurally related as well as to identify the different toxicological power of groups or substituents in such chemicals. On the other hand, the use of QSTR models for noncongeneric data sets permits the generalization of such mechanism to structurally diverse compounds as well as the identification of possible toxicophores of different structural nature.[17–20]

On the other hand, Markov models are well-known tools for characterizing biomolecules structure. Markov models have been used for analyzing biological sequence data and they have been used to find new genes from the open reading frames.[21,22] Another use of these models is data-based searching and multiple sequence alignment of protein families and protein domains. Protein turn types and sub-cellular locations have been successfully predicted.[23–26] Hubbard and Park[27] used amino acid sequence-based hidden Markov models to predict secondary structures. In this sense, Krogh et al.[23] have also proposed a hidden Markov model architecture. In addition, Markov's stochastic process has been used for protein folding recognition.[28] This approach can also be used for the prediction of protein signal sequences.[29,30] Another seminar works can be found related to the application of Markov chain theory to proteomic and bioinformatics. Chou applied Markov models to predict beta turns and their types, and the prediction of protein cleavage sites by HIV protease.[31–34] Anyhow, have not been very used Markov models to develop QSTR studies and predict drugs side effect.

In this connection, our group has introduced elsewhere a physically meaningful Markov model (Markovian Chemicals In Silico Design: MARCH-INSIDE) encoding molecular backbones information, with several applications in bioorganic medicinal chemistry. It allowed us to introduce matrix invariants such as stochastic entropies and spectral moments for the study of molecular properties. Specifically, the stochastic spectral moments introduced by our group have been largely used for small molecules QSAR problems including design of fluckicidal, anticancer and antihypertensive drugs. Applications to macromolecules have been restricted to the field of RNA without applications to proteins.[35–40] In addition, the entropy like molecular descriptors has demonstrated flexibility in many bioorganic and medicinal chemistry problems such as: estimation of anticoccidial activity, modeling the interaction between drugs and HIV-packaging-region RNA, and predicting proteins and virus activity.[41–46] In the field of QSTR our group has reported the first model to predict chemically-induced agranulocytocis by small-to-medium sized drug like molecules.[47]

However, in spite of several QSTR studies reported there have not been seriously studied almost drug side effects. Unfortunately, more than 1500 molecular descriptors reported have not only been applied to study drug side effects but have very disperse theoretical definition and sometimes not very well-established physical definition. Consequently, it becomes a forefront problem applying molecular descriptors to drugs side effect study but at the same time represent them in unified mathematical framework giving better opportunities for physicochemical interpretation.[48] In the current paper we attempt to develop a more serious physicochemical interpretation of the MARCH-INSIDE descriptors in thermodynamic terms, which allow us to contrast the relationship among these descriptors and topologic, flexibility, and quadratic molecular descriptors.[49] These new interpretations allow us to build up a molecular thermodynamic basis in free energy terms[50] for predicting how likely given drugs cause a specific side effect with respect to others side effects. This approach is able to take into consideration not only the molecular structure of the drug but the specific system the drug affects too. In particular it will be possible to correlate more than one property at a time, in our case, drugs side effects, making it superior against most of molecular descriptors, which simply permit to correlate no more than one property at a time, this advantage may be appropriately used in preliminary pharmacological or toxicological studies, especially for comparative studies in early stages of drug development. This study model a noncongeneric data set of 301 drugs of diverse molecular structure involved in 39 different side effects grouped in 11 affected systems, being 686 cases finally.

2. Methods

2.1. Markov thermodynamics for drug-target step-by-step interaction

We will consider a hypothetical situation in which a drug molecule is free in the space at an arbitrary initial time (t_0). It is then interesting to develop a simple stochastic model for a step-by-step interaction between the atoms of a drug molecule and a molecular receptor in the time on the induction of a side effect. For the sake of simplicity, we are going to consider from now on a general structureless receptor. Understanding as structureless molecular receptor a model of receptor which chemical structure it is not taken into consideration. The initial free energy of the drug–receptor interaction (${}^{0}g_j$) is a state function so a reversible process of interaction may be came apart on several elemental interactions between the j-th atom and the receptor.[50] Afterwards, interaction continues and we have to define the free energy of interaction between the j-th atom and the receptor given that i-th atom has been interacted at a previous time t_k (${}^{k}g_{ij}$). In particular, immediately after the first

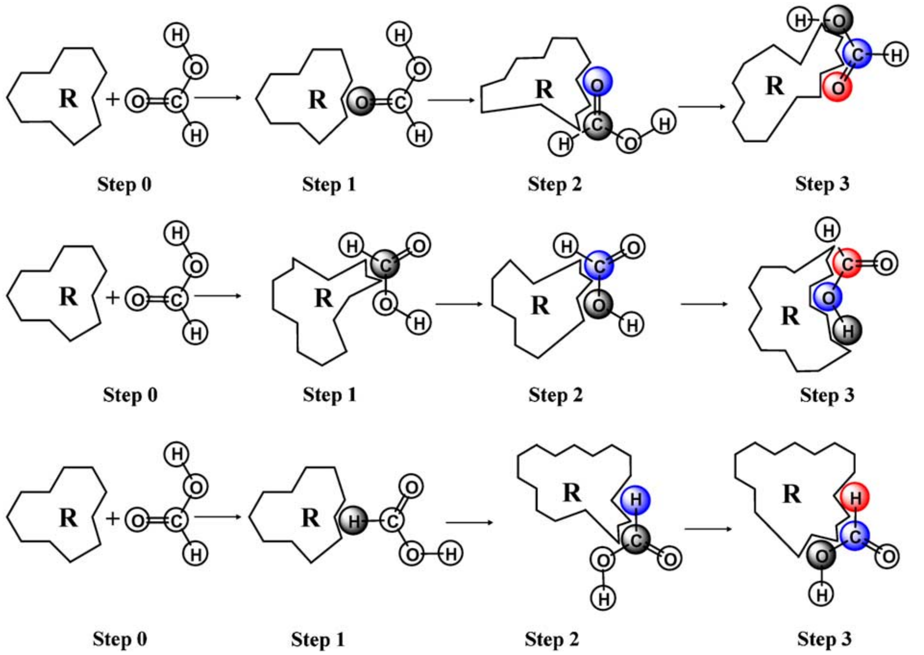


Figure 1. Stochastic drug-target step-by-step interaction.

interaction ($t_0 = 0$) takes place an interaction ${}^1g_{ij}$ at the time $t_1 = 1$ and so on. So, one can suppose that, atoms begin its interaction with the structureless molecular receptor binding to this receptor in discrete intervals of time t_k. However, there are several alternative ways in which such step-by-step binding processes may occur. Figure 1 illustrates this idea.

The free energy ${}^0g_j$ will be considered here as a function of the absolute temperature of the system and the equilibrium local constant of interaction between the j-th atom and the receptor (${}^0\Gamma_j$).[50] Additionally, the energy ${}^1g_{ij}$ can be defined by analogy as Γ_{ij}:

$$ {}^0g_j = R \cdot T \cdot \log {}^0\Gamma_j \tag{1} $$

$$ {}^1g_{ij} = R \cdot T \cdot \log {}^1\Gamma_{ij} \tag{2} $$

The present approach to drug–receptor interaction has two main drawbacks. The first is the difficulty on the definition of the constants. In this work, we solve the first question estimating k_j as the rate of occurrence (n_j) of the j-th atom on molecules inducing the effect under study by molecule–receptor interaction with respect to the number of atoms in the molecule (n). With respect to ${}^1\Gamma_{ij}$ we must take into consideration that once the j-th atom have interacted the preferred candidates for the next interaction are such i-th atoms bound to j by a chemical bond. Both constants can be then written down as:

$$ {}^0\Gamma_j = \frac{n_j}{n} = e^{\frac{{}^0g_j}{R\cdot T}} \tag{3} $$

$$ {}^1\Gamma_{ij} = \alpha_{ij} \cdot \frac{n_j}{n} = e^{\frac{{}^1g_{ij}}{R\cdot T}} \tag{4} $$

where, α_{ij} are the elements of the atom adjacency matrix, n_j, n, ${}^0g_j$, and ${}^1g_{ij}$ have been defined in the paragraph above, R is the gases constant, and T the absolute temperature. The second problem relates to the description of the interaction process at higher times $t_k > t_1$. Therefore, a Markov chain model (MC)[35–40] enables a simple calculation of the probabilities with which the drug–receptor interaction takes place in the time until the studied effect is achieved. In this work we are going to focus on drugs side effects. As depicted in Figure 1, this model deals with the calculation of the probabilities (${}^kp_{ij}$) with which any arbitrary molecular atom j-th bind to the structureless molecular receptor given that other atom i-th has been bound before; along discrete time periods t_k (k = 1,2,3,...); (k = 1 in gray), (k = 2 in blue), and (k = 3 in red) throughout the chemical bonding system.

The procedure described here considers as states of the MC the atoms of the molecule. The method arranges all the ${}^0\Gamma_j$ values in a vector (Γ) and all the ${}^1\Gamma_{ij}$ constants as a squared table of $n \times n$ dimension. After normalization of both the vector and the matrix we can built up the corresponding absolute initial probability vector φ and the stochastic matrix ${}^1\Pi$, which has the elements ${}^Ap_0(j)$ and ${}^1p_{ij}$, respectively. The elements ${}^Ap_0(j)$ of the above mentioned vector φ constitutes the absolute probabilities with which the j-th atom interact with the receptor at the initial time with respect to any atom in the molecule:

$$ {}^Ap_0(j) = \frac{{}^0\Gamma_j}{\sum_{a=1}^{m} {}^0\Gamma_a} = \frac{\frac{n_j}{n}}{\sum_{a=1}^{m} \frac{n_a}{n}} = \frac{\frac{1}{n} \cdot n_j}{\frac{1}{n} \cdot \sum_{a=1}^{m} n_a} = \frac{n_j}{\sum_{a=1}^{m} n_a} \tag{5} $$

where, m represents all the atoms in the molecule including the j-th, n_a is the rate of occurrence of any atom a including the *j*-th with value n_j. On the other hand, the matrix is called the one-step drug-target interaction stochastic matrix. $^1\Pi$ is built too as a squared table of order *n*, where *n* represents the number of atoms in the molecule. The elements ($^1p_{ij}$) of the one-step drug-target interaction stochastic matrix are the binding probabilities with which a *j*-th atom bind to a structureless molecular receptor given that other *i*-th atoms have been interacted before at time $t_1 = 1$ (considering $t_0 = 0$):

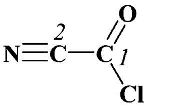
$$^1p_{ij} = \frac{^1\Gamma_{ij}}{\sum_{k=1}^{\delta+1} {^1\Gamma_{ik}}} = \frac{\alpha_{ij}\cdot\frac{n_j}{n}}{\sum_{k=1}^{\delta+1}\alpha_{ik}\cdot\frac{n_k}{n}} = \frac{\frac{1}{n}\cdot\alpha_{ij}\cdot n_j}{\frac{1}{n}\cdot\sum_{k=1}^{\delta+1}\alpha_{ik}\cdot n_k} = \frac{\alpha_{ij}\cdot n_{\mathrm{j}}}{\sum_{k=1}^{\delta+1}\alpha_{ik}\cdot n_k} \tag{6}$$

where, δ is the valence of the *j*-th atom. The calculation of φ and $^1\Pi$ is illustrated in Figure 2. By using, both φ and $^1\Pi$ and Chapman–Kolgomorov equations one can describe the further evolution of the system, determining the average constant of interaction between the *j*-th atom and the receptor at higher times. Summing up all the constants of interaction for each atom we can derive the stochastic molecular average constant of interaction ($^k\Gamma_{\mathrm{M}}$) between the drug and the receptor at a specific time:

$$^k\Gamma_{\mathrm{M}} = \varphi\cdot{^k\Pi}\cdot\Gamma = \varphi\cdot\left({^1\Pi}\right)^k\cdot\Gamma = \sum_{j=1}^{n}{^k\Gamma_j} = \sum_{j=1}^{n}{^A p_k(j)}\cdot{^0\Gamma_j} \tag{7}$$

Such a model is stochastic per se (probabilistic step-by-step atom–receptor interaction in time) but also considers molecular connectivity (the step-by-step atom union in space throughout the chemical bonding system). The selection of a Markov chain process[51,52] is not arbitrary. Due to atoms interactions are not dependent of previous atoms interactions we can affirm that a MCH-based model of a stochastic drug-target step-by-step interaction obeys perfectly to the main characteristics of

Structure:



Cyanoformyl chloride

$$\Pi^1_{Pan} = \begin{array}{c|ccccc} & \mathrm{O} & \mathrm{Br} & \mathrm{C_1} & \mathrm{C_2} & \mathrm{N} \\ \mathrm{O} & 5.775 & 0 & 27.302 & 0 & 0 \\ \mathrm{Br} & 0 & 25.566 & 14.58 & 0 & 0 \\ \mathrm{C_1} & 15.488 & 14.769 & 14.932 & 14.932 & 0 \\ \mathrm{C_2} & 0 & 0 & 13.467 & 13.467 & 10.55 \\ \mathrm{N} & 0 & 0 & 0 & 8.692 & 5.775 \end{array}$$

$$\Pi^1_{NoV} = \begin{array}{c|ccccc} & \mathrm{O} & \mathrm{Br} & \mathrm{C_1} & \mathrm{C_2} & \mathrm{N} \\ \mathrm{O} & 25.566 & 0 & 27.302 & 0 & 0 \\ \mathrm{Br} & 0 & 25.566 & 14.58 & 0 & 0 \\ \mathrm{C_1} & 63.712 & 76.434 & 65.448 & 65.448 & 0 \\ \mathrm{C_2} & 0 & 0 & 47.412 & 47.412 & 50.132 \\ \mathrm{N} & 0 & 0 & 0 & 22.846 & 25.566 \end{array}$$

$$\Pi^1_{DE} = \begin{array}{c|ccccc} & \mathrm{O} & \mathrm{Br} & \mathrm{C_1} & \mathrm{C_2} & \mathrm{N} \\ \mathrm{O} & 4.563 & 0 & 5.219 & 0 & 0 \\ \mathrm{Br} & 0 & 4.563 & 7.173 & 0 & 0 \\ \mathrm{C_1} & 14.299 & 12.345 & 14.955 & 14.955 & 0 \\ \mathrm{C_2} & 0 & 0 & 8.255 & 8.255 & 8.126 \\ \mathrm{N} & 0 & 0 & 0 & 4.692 & 4.563 \end{array}$$

Definition:

$$\Pi^1 = \begin{array}{c|ccccc} & \mathbf{O} & \mathbf{Cl} & \mathbf{C_1} & \mathbf{C_2} & \mathbf{N} \\ \mathbf{O} & \frac{O}{O+C_1} & 0 & \frac{C_1}{O+C_1} & 0 & 0 \\ \mathbf{Cl} & 0 & \frac{Cl}{C_1+Cl} & \frac{C_1}{C_1+Cl} & 0 & 0 \\ \mathbf{C_1} & \frac{O}{O+Cl+C_1+C_2} & \frac{Cl}{O+Cl+C_1+C_2} & \frac{C_1}{O+Cl+C_1+C_2} & \frac{C_2}{O+Cl+C_1+C_2} & 0 \\ \mathbf{C_2} & 0 & 0 & \frac{C_1}{C_1+C_2+N} & \frac{C_2}{C_1+C_2+N} & \frac{N}{C_1+C_2+N} \\ \mathbf{N} & 0 & 0 & 0 & \frac{C_2}{C_2+N} & \frac{N}{C_2+N} \end{array}$$

Figure 2. Definition and calculation of Π^1 matrix for a specific compound in three particular cases of side effects. The element symbol is used to denote the value of the rate of recurrence [i.e., Cl represents the rate of recurrence (n_{Cl}) of chlorine atom for the specific side effect].

MCH (a memoryless property). This implies that the probability of the occurrence of an event (atom union) does not depend on the history of the process. In other words, such a model will not depend on previous atoms union.

2.2. Statistical analysis

As a continuation of the previous sections, we can attempt to develop a simple linear QSAR using the MARCH-INSIDE methodology, as defined previously, with the general formula:

$$\begin{aligned}\mathrm{SE}_x = b + b_0 \cdot {}^0\Gamma_\mathrm{M} + b_1 \cdot {}^1\Gamma_\mathrm{M} + b_2 \cdot {}^2\Gamma_\mathrm{M} \\ + b_3 \cdot {}^3\Gamma_\mathrm{M} \cdots + b_k \cdot {}^k\Gamma_\mathrm{M}\end{aligned} \tag{8}$$

Here, ${}^k\Gamma_\mathrm{M}$ act as the molecular descriptors. We selected linear discriminant analysis (LDA)[53,54] to fit the classification functions. The model deals with the classification of a set of compounds with diverse side effects. A dummy variable (SE_x) codifies the side effect studied. This variable indicates either the presence ($\mathrm{SE}_x = 1$) or absence ($\mathrm{SE}_x = -1$) of the side effect studied. In Eq. 8, b_k represents the coefficients of the classification function, determined by the least square method as implemented in the LDA module of the STATISTICA 6.0 software package.[55] Forward stepwise was fixed as the strategy for variable selection.[53,54]

The quality of LDA models was determined by examining Wilk's *U* statistic, Fisher ratio (*F*), and the *p*-level (*p*). We also inspected the percentage of good classification and the ratios between the cases and variables in the equation and variables to be explored in order to avoid over-fitting or chance correlation. Validation of the model was corroborated by re-substitution of cases in four predicting series.[37]

Clustering of compounds was carried out after previous perform of a canonical analysis using the algorithms implemented in the advanced options for LDA in the STATISTICA 6.0. This analysis offers as outputs the scores of every case for successive canonical roots, which are orthogonal centered equations explaining decreased amounts of variance. Consequently we can plot the scores for each compound in a Cartesian system of coordinates and using a symbol code visually exploring the possibility of clusters formations.[37]

2.3. Data set methodology

The data set was conformed by a set of more frequently used drugs, which produce side effects in different human organs systems, being these ones extensively tested in clinic and the side effects reported obtained by pharmacovigilance studies. The use of marketed drugs in data set confers a high confidence about the side effect reported. The set of drugs where extracted from a report of drugs side effects listed in literature.[56] The data set was conformed by 39 different drugs side effects grouped in 11 affected systems for 301 drugs, being 686 cases finally, taking into consideration that all side effects groups were statistically represented having each one at least seven drugs in order to perform a balanced training series.

2.4. Laboratory animals and biological assay

Sufficient quantities of analytical grade G-1 for biological assays were purchased from the Chemicals Bioactive Center. Differential counting of limphocytes was carried out as recommended in the literature.[57] Balb/c mice were selected as a biological model.[58] Healthy Balb/c mice of both sexes were purchased, along with food, from the 'Centro Nacional de Animales de Laboratorio (CENPALAB)', Cuba. Quarantine, labeling, acclimatization, and good maintenance conditions of animals were strictly adhered.[59]

3. Results and discussion

3.1. Model

Eq. 7 constitutes in mathematical terms a vector–matrix–vector form. Panoply of these transformations has been previously used in QSAR studies for a long time. For instance, the first molecular descriptor defined in a chemical context the Wiener index *W* (Eq. 9) is a quadratic form.[60] In addition, several other classic Zagreb indices M_1 (10) and M_2 (11), Harary number *H* (12), Randic invariant χ (13), valence connectivity index χ^v (14), the Balaban index *J* (15), the MTI index (16), the global flexibility index GS (17), the Moreau–Boroto autocorrelation ATS_d (18), and the Cluj–Detour index CJD (19), just to mention a few examples, may be expressed all of them quadratic forms. The representation of all these indices was recently unified in the literature.[61] More recently other topologic indices based on quadratic forms as the so called quadratic indices $q_k(X)$ (20) have been introduced by our group as well:[62]

$$W = \frac{1}{2}\left(u \cdot D \cdot u^\mathrm{T}\right) \tag{9}$$

$$M_1 = v \cdot A \cdot u^\mathrm{T} \tag{10}$$

$$M_2 = \frac{1}{2}\left(v \cdot A \cdot v^\mathrm{T}\right) \tag{11}$$

$$H = \frac{1}{2}\left(u \cdot D^{-k} \cdot u^\mathrm{T}\right) \tag{12}$$

$$\chi = v' \cdot A \cdot v'^\mathrm{T} \tag{13}$$

$$\chi^v = v'' \cdot A \cdot v''^\mathrm{T} \tag{14}$$

$$J = \frac{1}{2} \cdot C \cdot \left(d' \cdot A \cdot d'^\mathrm{T}\right) \tag{15}$$

$$MTI = v \cdot (A + D) \cdot u^{\mathrm{T}} \tag{16}$$

$$GS = \frac{2}{A \cdot (A-1)} \cdot \left(u \cdot L \cdot u^{\mathrm{T}}\right) \tag{17}$$

$$ATS_d = w^{\mathrm{T}} \cdot {}^{m}B \cdot w \tag{18}$$

$$CJD = \frac{1}{2} u^{\mathrm{T}} \cdot \Delta \cdot u \tag{19}$$

$$q_k(X) = x^{\mathrm{T}} \cdot M \cdot x \tag{20}$$

where, D, A, D^{-k}, ${}^{m}B$, Δ, L, and M are matrices related to distance, atom adjacency, sparse, Laplace, pseudograph matrices, and others. On the other hand, u, v, v', v'', w, x, and d are vectors related to unitary, vertex degree, Randic atom degree, valence degree, atom weight, atom electronegativity, and distance among others. All the vectors and matrices used in expressions (9)–(20) have been exhaustively explained in the literature reported and references therein cited, see therein for details.

In the present work we propose to call all these molecular indices the deterministic vector–matrix–vector forms by opposition to our stochastic forms. The main advantage of the present stochastic forms is the possibility of deriving average thermodynamic parameters depending on the probability of the states of the MC, which fit on more clearly physicochemical sense with respect to classic quadratic forms. In specific, this work introduces for the first time a Markov form to calculate thermodynamic parameters of the drug-target interaction process considering in a unified scheme: time, chemical structure, and system including drug side effects.

Another advantage of these vector–matrix–vector forms constitute the fact that it was not necessary considering different rates of occurrence for atoms of the same element but having different configuration, for example: sp^3, sp^2, and sp carbons all were considered with the same rate of occurrence for a specific side effect, the rate of carbon atoms. It was possible due to the p_{ij} values clearly distinguished among these atoms because of the different connectivity (see Fig. 2). It is clear from Figure 2 that atoms with different connectivity or configuration will have a different probability of union to the structure-less molecular receptor in spite of having the same rate of occurrence.

Once we perform a representative and balanced training series selection it could be used to fit the classification functions. The models where subjected to the principle of parsimony. Then, we chose a function with high statistical significance but having few parameters ($b_k\ {}^{k}\Gamma_{\mathrm{M}}$) as possible to each of 39 studied side effects.

In order to derive a classification function that permits the classification of drugs as positive (presence of side effect) or negative (absence of side effect) we use the LDA in which stochastic molecular average constant of interaction (${}^{k}\Gamma_{\mathrm{M}}$) are used as independent variables. The classification models obtained to each studied side effect are given below in Table 1 together with the statistical parameters of the LDA, validations of the current model by re-substitution of cases in four predicting

Table 1. Overall train accuracy-validation (CV) predictability, and models for different drugs side effects

Side effects	Train	CV	Model
Breathing manifestations			
Infiltrated lung (IL)	92.3	98.1	$\mathrm{IL} = -87.38 + 35.17{}^{0}\Gamma_{\mathrm{M}}$
Bronchospasm (Brch)	100	100	$\mathrm{Brch} = -176.69 + 50.12{}^{0}\Gamma_{\mathrm{M}}$
Total	**97.1**	**97.5**	$N =$ **35**, $U =$ **0.114**, $F =$ **255.7**
Cardiovascular manifestations			
Exacerbations of angina pectoris (EAP)	100	100	$\mathrm{EAP} = -112.70 + 59.90{}^{0}\Gamma_{\mathrm{M}} - 1.421{}^{4}\Gamma_{\mathrm{M}} + 0.689{}^{1}\Gamma_{\mathrm{M}}$
Arrhythmias (Arr)	100	100	$\mathrm{Arr} = -1862.78 + 245.41{}^{0}\Gamma_{\mathrm{M}} - 5.67{}^{4}\Gamma_{\mathrm{M}} + 2.74{}^{1}\Gamma_{\mathrm{M}}$
Edema or liquid retention in heart inadequacy (ELRHI)	100	98.9	$\mathrm{ELRHI} = -594.65 + 138.47{}^{0}\Gamma_{\mathrm{M}} - 3.21{}^{4}\Gamma_{\mathrm{M}} + 1.56{}^{1}\Gamma_{\mathrm{M}}$
Hipertensión (HyperT)	100	100	$\mathrm{HyperT} = -137.72 + 66.37{}^{0}\Gamma_{\mathrm{M}} - 1.60{}^{4}\Gamma_{\mathrm{M}} + 0.77{}^{1}\Gamma_{\mathrm{M}}$
Hypotension (HypoT)	86.4	96.6	$\mathrm{HypoT} = -411.95 + 115.20{}^{0}\Gamma_{\mathrm{M}} - 2.55{}^{4}\Gamma_{\mathrm{M}} + 1.23{}^{1}\Gamma_{\mathrm{M}}$
Thromboembolism (Thr)	100	100	$\mathrm{Thr} = -315.70 + 100.72{}^{0}\Gamma_{\mathrm{M}} - 2.45{}^{4}\Gamma_{\mathrm{M}} + 1.19{}^{1}\Gamma_{\mathrm{M}}$
Total	**97.6**	**96.8**	$N =$ **125**, $U =$ **0.003**, $F =$ **158.6**
Hematological manifestations			
Agranulocytosis (Agr)	85	98.8	$\mathrm{Agr} = -391.15 + 149.73{}^{0}\Gamma_{\mathrm{M}} - 12.99{}^{2}\Gamma_{\mathrm{M}} + 9.39{}^{1}\Gamma_{\mathrm{M}}$
Hemolytic anemia (HA)	100	100	$\mathrm{HA} = -903.99 + 227.83{}^{0}\Gamma_{\mathrm{M}} - 19.63{}^{2}\Gamma_{\mathrm{M}} + 14.20{}^{1}\Gamma_{\mathrm{M}}$
Hemolytic anemia (in deficit of G6PD) (HAG6PD)	100	100	$\mathrm{HAG6}PD = -162.71 + 96.29{}^{0}\Gamma_{\mathrm{M}} - 8.28{}^{2}\Gamma_{\mathrm{M}} + 5.98{}^{1}\Gamma_{\mathrm{M}}$
Pancytopenia (Pan)	90	100	$\mathrm{Pan} = -334.62 + 138.43{}^{0}\Gamma_{\mathrm{M}} - 11.85{}^{2}\Gamma_{\mathrm{M}} + 8.57{}^{1}\Gamma_{\mathrm{M}}$
Platelet aggregation alterations (PAA)	100	100	$\mathrm{PAA} = -640.64 + 191.50{}^{0}\Gamma_{\mathrm{M}} - 16.22{}^{2}\Gamma_{\mathrm{M}} + 11.74{}^{1}\Gamma_{\mathrm{M}}$
Total	**95.1**	**99.1**	$N =$ **103**, $U =$ **0.012**, $F =$ **90.5**

Table 1 (*continued*)

Side effects	Train	CV	Model
Gastrointestinal manifestations			
Constipation or ileo (CoI)	90.9	100	CoI = $-41.91 + 20.78\,{}^{0}\Gamma_{\mathrm{M}} + 0.33\,{}^{2}\Gamma_{\mathrm{M}} - 0.39\,{}^{3}\Gamma_{\mathrm{M}}$
Diarrhea or colitis (DoC)	100	100	DoC = $-208.83 + 47.26\,{}^{0}\Gamma_{\mathrm{M}} + 0.60\,{}^{2}\Gamma_{\mathrm{M}} - 0.70\,{}^{3}\Gamma_{\mathrm{M}}$
Diffuse hepatocellular Damage (DHD)	95	98.8	DHD = $-93.19 + 31.41\,{}^{0}\Gamma_{\mathrm{M}} + 0.63\,{}^{2}\Gamma_{\mathrm{M}} - 0.73\,{}^{3}\Gamma_{\mathrm{M}}$
Mouth dryness (MD)	86.7	100	MD = $-65.40 + 26.19\,{}^{0}\Gamma_{\mathrm{M}} + 0.62\,{}^{2}\Gamma_{\mathrm{M}} - 0.72\,{}^{3}\Gamma_{\mathrm{M}}$
Nausea or vomit (NoV)	100	100	NoV = $-1562.42 + 129.82\,{}^{0}\Gamma_{\mathrm{M}} + 2.52\,{}^{2}\Gamma_{\mathrm{M}} - 2.95\,{}^{3}\Gamma_{\mathrm{M}}$
Pancreatitis (Pat)	100	100	Pat = $-53.90 + 23.70\,{}^{0}\Gamma_{\mathrm{M}} + 0.46\,{}^{2}\Gamma_{\mathrm{M}} - 0.53\,{}^{3}\Gamma_{\mathrm{M}}$
Peptic or hemorrhagic ulceration (PoHU)	93.8	100	PoHU = $-63.21 + 25.70\,{}^{0}\Gamma_{\mathrm{M}} + 0.20\,{}^{2}\Gamma_{\mathrm{M}} - 0.24\,{}^{3}\Gamma_{\mathrm{M}}$
Total	**97**	**99.4**	N = **164**, U = **0.002**, F = **190.9**
Dermal manifestations			
Acne (Ac)	50	90	Ac = $-212.11 + 117.62\,{}^{0}\Gamma_{\mathrm{M}}$
Alopecia (Alp)	100	100	Alp = $-1456.17 + 309.04\,{}^{0}\Gamma_{\mathrm{M}}$
Diverse erythema (DE)	91.7	100	DE = $-196.81 + 113.27\,{}^{0}\Gamma_{\mathrm{M}}$
Photodermatitis (PhD)	100	100	PhD = $-268.05 + 132.31\,{}^{0}\Gamma_{\mathrm{M}}$
Total	**88**	**92.3**	N = **50**, U = **0.004**, F = **3920**
Systemic phenomena			
Anaphylaxis (Anph)	100	100	Anph = $-46.36 + 17.29\,{}^{0}\Gamma_{\mathrm{M}}$
Lupus erythematosus (LE)	100	100	LE = $-14.43 + 9.39\,{}^{0}\Gamma_{\mathrm{M}}$
Fever (Fv)	100	100	Fv = $-133.28 + 29.55\,{}^{0}\Gamma_{\mathrm{M}}$
Total	**100**	**100**	N = **47**, U = **0.042**, F = **506.1**
Endocrine manifestations			
Galactorrhea (amenorrhea) (Gal)	100	100	Gal = $-71.25 + 41.25\,{}^{0}\Gamma_{\mathrm{M}}$
Livid decrease and impotence (LDI)	100	100	LDI = $-98.16 + 48.52\,{}^{0}\Gamma_{\mathrm{M}}$
Thyroid function test disorders (TFTD)	88.9	94.4	TFTD = $-30.62 + 26.76\,{}^{0}\Gamma_{\mathrm{M}}$
Total	**97**	**95**	N = **33**, U = **0.12**, F = **110**
Metabolic manifestations			
Hyperglycemia (HyperG)	100	100	HyperG = $-56.91 + 48.06\,{}^{0}\Gamma_{\mathrm{M}}$
Hypopotassemia (HypoP)	100	100	HypoP = $-141.22 + 75.92\,{}^{0}\Gamma_{\mathrm{M}}$
Total	**100**	**100**	N = **18**, U = **0.086**, F = **170.8**
Neurological manifestations			
Convulsions (Cvs)	100	100	Cvs = $-597.82 + 166.92\,{}^{0}\Gamma_{\mathrm{M}}$
Extrapyramidals effects (EE)	100	100	EE = $-252.34 + 108.36\,{}^{0}\Gamma_{\mathrm{M}}$
Total	**100**	**100**	N = **33**, U = **0.026**, F = 1158.1
Psychiatric manifestations			
Deliriums orconfusional states (DoCS)	100	100	DoCS = $-211.26 + 70.9\,{}^{0}\Gamma_{\mathrm{M}} + 0.79\,{}^{1}\Gamma_{\mathrm{M}} - 1.46\,{}^{5}\Gamma_{\mathrm{M}}$
Dysfunctions of the dream (DD)	100	100	DD = $-84.02 + 44.48\,{}^{0}\Gamma_{\mathrm{M}} + 0.28\,{}^{1}\Gamma_{\mathrm{M}} - 0.46\,{}^{5}\Gamma_{\mathrm{M}}$
Somnolence (Snl)	90.9	96.6	Snl = $-270.26 + 80.03\,{}^{0}\Gamma_{\mathrm{M}} + 0.41\,{}^{1}\Gamma_{\mathrm{M}} - 0.60\,{}^{5}\Gamma_{\mathrm{M}}$
Total	**96.4**	**94.6**	N = **55**, U = **0.04**, F = **66.5**
Muscular-sΓeletal manifestations			
Myopathy or myalgia (MoM)	100	100	MoM = $-276.33 + 92.77\,{}^{0}\Gamma_{\mathrm{M}} - 0.49\,{}^{5}\Gamma_{\mathrm{M}}$
Osteoporosis (Ost)	100	100	Ost = $-61.14 + 43.41\,{}^{0}\Gamma_{\mathrm{M}} - 0.23\,{}^{5}\Gamma_{\mathrm{M}}$
Total	**100**	**100**	N = **23**, U = **0.027**, F = **361.4**

series results and percents of good classification to each model.

In the models the coefficient U is the Wilk's statistics (statistic for the overall discrimination is computed as the ratio of the determinant of the within-groups variance/covariance matrix over the determinant of the total variance/covariance matrix) and F is the Fisher ratio. The Wilk's U-statistic is the standard statistic that is used to denote the statistical significance of the discriminatory power of the current model.[35,63] U-statistic values range from 1.0 (no discriminatory power) to 0.0 (perfect discriminatory power).

In order to simplify the equations for the purposes of interpretation and the possibility of graphical representation, we performed a canonical analysis[64] for two systems of side effects with the only purpose to illustrate the capability of the equations obtained to condense more than two side effects groups in only one simple equation (root function) and its ability to discriminate between several side effects groups.

For the cardiovascular manifestations side effects group the main root obtained (root 1) proved to be a simple equation centered to 0:

$$\begin{aligned}\text{Root } 1 &= -1.0995\,{}^{0}\Gamma_{\mathrm{M}} - 3.4955\,{}^{1}\Gamma_{\mathrm{M}} + 3.8155\,{}^{4}\Gamma_{\mathrm{M}}\\ \text{Root } 2 &= -0.0180\,{}^{0}\Gamma_{\mathrm{M}} + 11.8134\,{}^{1}\Gamma_{\mathrm{M}} - 11.1715\,{}^{4}\Gamma_{\mathrm{M}}\\ \text{Root } 3 &= -0.06376\,{}^{0}\Gamma_{\mathrm{M}} - 9.344638\,{}^{1}\Gamma_{\mathrm{M}} + 10.13876\,{}^{4}\Gamma_{\mathrm{M}}\end{aligned} \tag{21}$$

This canonical root presented an eigen value of 329.37 and a high regression coefficient of 0.9985, which it is statistically significant (p-level > 0.05) explaining the 99.69% of the variance of the cases in the used data, together with a Chi-squared statistic of 700.43. Figure 3 shows six side effects groups clearly distinguished and discriminated in a canonical space graph.

The psychiatric manifestations side effects group was visibly distinguished also by root 1 (main root obtained):

$$\begin{aligned}\text{Root } 1 &= -0.99657\,{}^{0}\Gamma_{\mathrm{M}} - 0.55270\,{}^{1}\Gamma_{\mathrm{M}} + 0.37741\,{}^{5}\Gamma_{\mathrm{M}}\\ \text{Root } 2 &= -0.0683\,{}^{0}\Gamma_{\mathrm{M}} - 10.3341\,{}^{1}\Gamma_{\mathrm{M}} + 10.1294\,{}^{5}\Gamma_{\mathrm{M}}\end{aligned} \tag{22}$$

This canonical root presented an eigen value of 17.93 and a high regression coefficient of 0.9732, which is statistically significant (p-level > 0.05) explaining the 94.72% of the variance of the cases in the used data, together with a Chi-squared statistic of 163.90. Figure 4 shows three side effects groups visibly distinguished and discriminated in a canonical space graph. In both cases, root 2 may be used to discriminate the cases within each group.

Aimed at finding some similarity with others descriptors we could contrast our stochastic vector–matrix–vector forms (${}^{k}\Gamma_{\mathrm{M}}$) with Toporov optimization of correlation weights of local graph invariants named flexible descriptors[65–69] (do not confuse with flexibility descriptors). In this sense, both descriptors take into consideration more than one parameter. In flexible descriptors case, it is taken into consideration the abstract parameters (weights), which can be optimized in function of the pursued objectives. On the other hand, the parameters our molecular descriptors take into consideration cannot be optimized, but have a direct physicochemical interpretation, such aspects have been analyzed in previous paragraphs of the methods section.

3.2. Experimental corroboration of some theoretic predictions for G-1 (2-bromo-5-[2-bromo-2-nitrovinyl] furan)

Finally, in order to exemplify the use of the classification functions obtained we decided to use G-1 in testing experimentally the ability to induce a specific side effect (pancytopenia). Maximum changes in blood, particularly the depletion of lymphocytes related with the dose is a fundamental point to arrive to conclusions in this assay. Experiments carried out in mice using as criteria the lymphocytes count provide evidences of occurrence of pancytopenia, reversible in a short time in groups treated with G-1 (as compared to control groups) when the time of administration and doses was varied (see Table 2) a very important parameter for the detection of drug induced blood dyscracias.[70,71] Statistically significant differences in the lymphocytes count could be detected after treatment with different doses of G-1 at

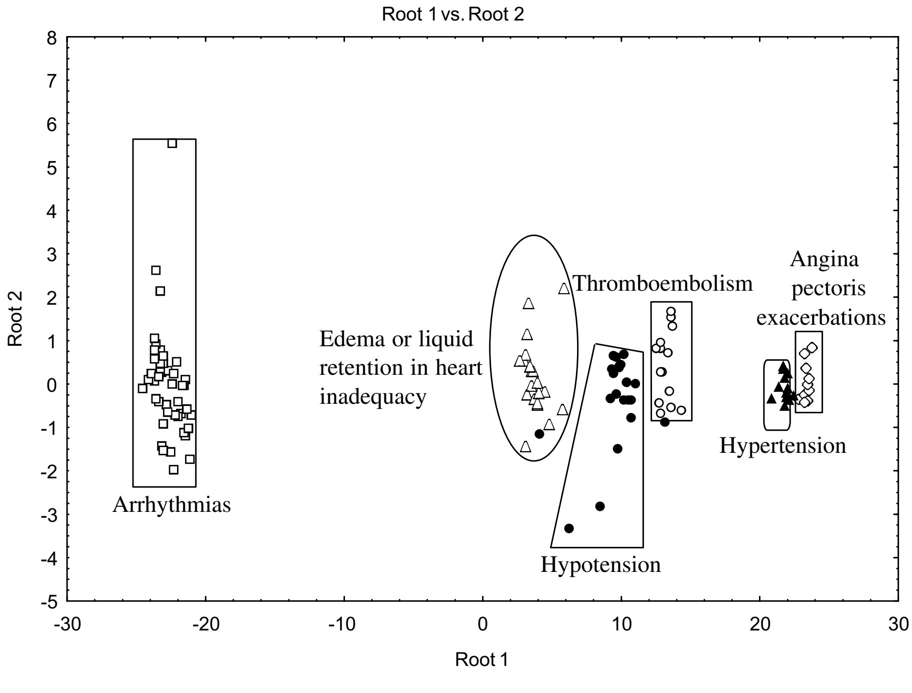


Figure 3. Canonical roots analysis for cardiovascular manifestations.

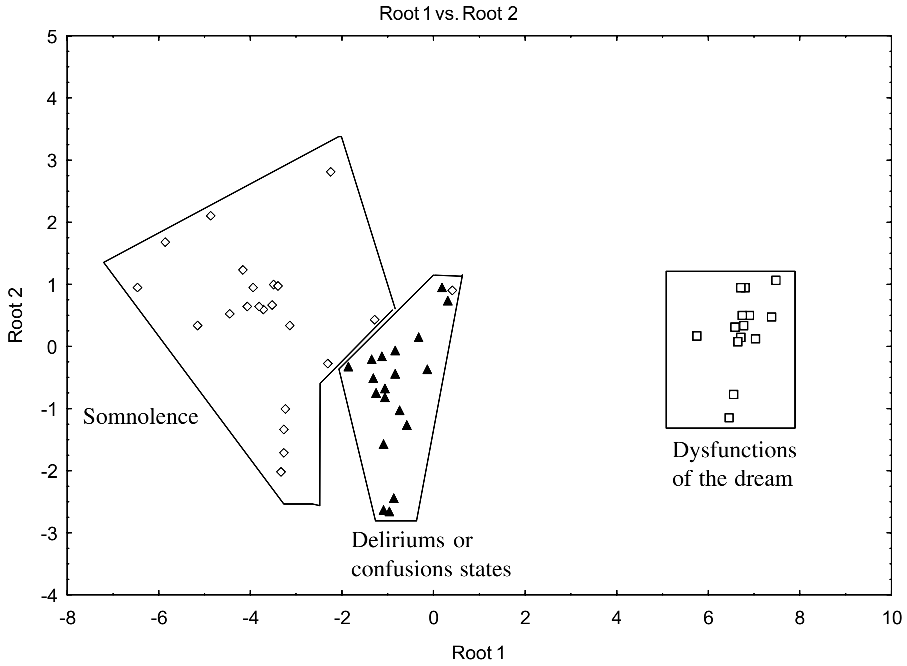
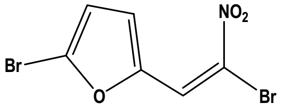


Figure 4. Canonical roots analysis for psychiatric manifestations.

Table 2. $^{k}\Gamma_{M}$ Calculation, hematological manifestations theoretical predictions, and differential count of lymphocytes at different times and doses in mouse for G-1, vehicle, positive control, and negative control solutions



Theoretical predictions for G-1						
Hematological manifestations	Molecular indices for G-1					
	$^{0}\Gamma_{M}$	$^{1}\Gamma_{M}$	$^{2}\Gamma_{M}$	Obs. Class.[a]	Pred. Class.[a]	Probability
Hemolytic anemia				—	1	1.0000
(in deficit of G6DP) (HAG6PD)	3.00	15.39	11.62			
Hemolytic anemia (HA)	7.11	35.77	26.81	—	0	0.0001
Platelet aggregation				—	0	0.0606
Alterations (PAA)	5.98	31.37	23.75			
Agranulocytosis (Agr)	4.20	20.76	15.61	—	0	0.0001
Pancytopenia (Pan)	**4.17**	**20.70**	**15.70**	**1**	**1**	**0.7491**
Experimental corroboration (control groups)						
Time (h)				48	72	96
Doses (mg/kg)					ΔU/L%[c]	
Migliol (0 mg/kg of G1)				0.78 ± 0.077	0.76 ± 0.04	0.74 ± 0.077
Negative (0 mg/kg of G1)				0.78 ± 0.04	0.78 ± 0.04	0.78 ± 0.04
Experimental corroboration (groups treated with G-1)						
Time (h)				48	72	96
Doses (mg/kg)					Δ U/L%[b]	
185.6				0.62 ± 0.02	0.75 ± 0.05	0.78 ± 0.02
61.8				0.61 ± 0.27	0.78 ± 0.02	0.85 ± 0.27
23.4				0.64 ± 0.19	0.80 ± 0.07	0.81 ± 0.07
12.3				0.69 ± 0.09	0.78 ± 0.01	0.75 ± 0.08
9.8				0.72 ± 0.1	0.71 ± 0.12	0.82 ± 0.03

[a] The value is 1 when it was experimentally corroborated the specific side effect for G-1; 0 when G-1 it was not detected such effect, and—when it has not been experimentally studied yet.

[b] Difference (arithmetic mean for the respective group) between the lymphocytes count after treatment U/L (after) with respect to lymphocytes count before the treatment U/L (before) in units (cells) per liter, that is, ΔU/L% = [U/L (after) − U/L (before)] × 100.

the 0.05 p-level by means of a dependent pairwise student test.[59] Additionally, the posterior probability predicted for G-1 ($P\% = 74.91$) coincides with the experimental results. An example of theoretical predictions and ${}^{k}\Gamma_{M}$ calculations of G-1 for a specific system is shown in Table 2.

4. Concluding remarks

The fusion of high throughput screening and QSAR/QSTR[35–47] techniques in attempt to minimize the costs in terms of time, financial, human, and animal resources is becoming a viable alternative to massive screening. The results described here (high percentages of good classification in training and predicting set as well as experimental corroboration results) have demonstrated that MARCH-INSIDE methodology encode molecular backbones information, with several applications in bioorganic medicinal chemistry. Specifically, stochastic molecular average constant of interaction (${}^{k}\Gamma_{M}$) is able to provide a physicochemical direct interpretation for drug-target step-by-step interaction taking into consideration not only the molecular structure of the drug but the specific system the drug affects too. In particular, thru this molecular descriptor will be possible correlate more than one property at time (in our case, drugs side effects) having a more serious physicochemical interpretation in thermodynamic terms. This fact makes the present descriptors superior weigh against most of molecular descriptors, which correlate no more than one property at a time.[72] This advantage may be appropriately used in preliminary pharmacological or toxicological studies, especially for comparative studies in drug development early stages.

Acknowledgements

M.Cz-M. and H.G.-D. would like to express their most sincere gratitude to Prof. Castro, E. and Prof. Toropov, A. A. for kindly supplying some very interesting materials on the topic of flexible indices. All the authors acknowledge Prof. Chi-Huey Wong, Editor of BMC for prompt attention and kind suggestions made, which contribute to increase the quality of the present paper.

Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.bmc.2004.11.030.

References and notes

1. Lutz, M. W.; Menius, J. A.; Laskody, R. G.; Domanico, P. L.; Goetz, A. G.; Saussy, D. L.; Rimele, T. *Network Sci.* **1996**, *2*(9), September.
2. Loew, G. H.; Villar, H. O.; Alkorta, Y. *Pharm. Res.* **1993**, *10*, 475.
3. Briggs, J. M.; Marrone, T. J.; McCammon, J. A. *Trends Cardiovasc. Med.* **1996**, *6*, 529.
4. Wess, G. *Drug Discov. Today* **1996**, *1*, 529.
5. Cronin, M. T. D. *Pharm. Pharmacol. Commun.* **1998**, 157–163.
6. Lewis, D. E. V. Computer Assisted Methods in the Evaluation of Chemical Toxicity. In *Reviews in Computational Chemistry*; Lipkowitz, K. B., Boyd, D. B., Eds.; VCH: New York, 1992; Vol. 3, pp 173–222.
7. Cronin, M. T. D.; Dearden, J. C. *Quant. Struct.-Act. Relat.* **1995**, *4*, 1–7.
8. Cronin, M. T. D.; Dearden, J. C. *Quant. Struct.-Act. Relat.* **1995**, *4*, 117–120.
9. Cronin, M. T. D.; Dearden, J. C. *Quant. Struct.-Act. Relat.* **1995**, *4*, 329–334.
10. Cronin, M. T. D.; Dearden, J. C. *Quant. Struct.-Act. Relat.* **1995**, *4*, 518–523.
11. Dearden, J. C.; Cronin, M. T. D.; Dobbs, A. J. *Chemosphere* **1995**, *31*, 2521–2528.
12. Roberts, D. W. An Analysis of Published Data on Fish Toxicity of Nitrobenzenes and Aniline Derivatives. In *QSAR in Environmental Toxicology-II*; Kaiser, K. L. E., Ed.; D. Reidel: Dordrecht, The Netherlands, 1987; pp 295–308.
13. Dearden, J. C.; Cronin, M. T. D.; Schultz, T. W.; Lin, D. T. *Quant. Struct.-Act. Relat.* **1995**, *4*, 427–432.
14. Debnath, A. K.; de Compadre, R. L. L.; Debnath, G.; Shusterman, A. J.; Hansch, C. *J. Med. Chem.* **1991**, *4*, 427–432.
15. Roberts, D. W. *Chem. Res. Toxicol.* **1995**, *8*, 545–551.
16. Mekenyan, O.; Roberts, D. W.; Karcher, W. *Chem. Res. Toxicol.* **1997**, *10*, 994–1000.
17. González, M. P.; Morales, A. H.; Molina, R. *Polymer* **2004**, *45*, 2773.
18. González, M. P.; Morales, A. H.; González-Díaz, H. *Polymer* **2004**, *45*, 2073.
19. Morales, A. H.; González, M. P.; Rieumont, J. B. *Polymer* **2004**, *45*, 2045.
20. González, M. P.; González-Díaz, H.; Cabrera-Pérez, M. A.; Molina, R. R. *Bioorg. Med. Chem.* **2004**, *12*, 735.
21. Vorodovsky, M.; Koonin, E. V.; Rudd, K. E. *Trends Biochem. Sci.* **1994**, *19*, 309.
22. Vorodovsky, M.; Macininch, J. D.; Koonin, E. V.; Rudd, K. E.; Médigue, C.; Danchin, A. *Nucl. Acid Res.* **1995**, *23*, 3554.
23. Krogh, A.; Brown, M.; Mian, I. S.; Sjeander, K.; Haussler, D. *J. Mol. Biol.* **1994**, *235*, 1501.
24. Chou, K.-C. *Biopolymer* **1997**, *42*, 837.
25. Yuan, Z. *FEBS Lett.* **1999**, *451*, 23.
26. Hua, S.; Sun, Z. *Bioinformatics* **2001**, *17*, 721.
27. Hubbard, T. J.; Park, J. *Proteins Struct. Funct. Genet.* **1995**, *23*, 398.
28. Di Francesco, V.; Munson, P. J.; Garnier, J. *Bioinformatics* **1999**, *15*, 131.
29. Chou, K.-C. *Curr. Protein Pept. Sci.* **2002**, *3*, 615.
30. Chou, K.-C. *Peptides* **2001**, *22*, 1973.
31. Chou, K.-C. *Anal. Biochem.* **2000**, *286*, 1.
32. Chou, K.-C. *J. Biol. Chem.* **1993**, *268*, 16938.
33. Chou, K.-C. *Anal. Biochem.* **1996**, *233*, 1.
34. Chou, K.-C.; Zhang, C. T. *J. Protein Chem.* **1993**, *12*, 709.
35. Gonázlez-Díaz, H.; Olazábal, E.; Castañedo, N.; Hernádez, S. I.; Morales, A.; Serrano, H. S.; Gonzlez, J.; Ramos de, A. R. *J. Mol. Mod.* **2002**, *8*, 237.
36. González-Díaz, H.; Gia, O.; Uriarte, E.; Hernádez, I.; Ramos, R.; Chaviano, M.; Seijo, S.; Castillo, J. A.; Morales, L.; Santana, L.; Akpaloo, D.; Molina, E.; Cruz, M.; Torres, L. A.; Cabrera, M. A. *J. Mol. Mod.* **2003**, *9*, 395.

37. González-Díaz, H.; Hernández, S. I.; Uriarte, E.; Santana, L. *Comput. Biol. Chem.* **2003**, *27*, 217.
38. González-Díaz, H.; Ramos de, A. R.; Molina, R. R. *Bull. Math. Biol.* **2003**, *65*, 991.
39. González-Díaz, H.; Ramos de, A. R.; Uriarte, E. *Online J. Bioinf.* **2002**, *1*, 83.
40. González-Díaz, H.; Uriarte, E.; Ramos de A. R. *Bioorg. Med. Chem.*, in press, see doi:10.1016/j.bmc.2004.10.024.
41. González-Díaz, H.; Molina, R. R.; Uriarte, E. *Bioorg. Med. Chem. Lett.* **2004**, *14*, 4691–4695.
42. González-Díaz, H.; Ramos de, A. R.; Molina, R. R. *Bioinformatics* **2003**, *19*, 2079.
43. González-Díaz, H.; Molina, R. R.; Uriarte, E. *Polymer* **2004**, *45*, 3845.
44. Ramos de, A. R.; González-Díaz, H.; Molina, R.; González, M. P.; Uriarte, E. *Bioorg. Med. Chem.* **2004**, *12*, 4815.
45. Ramos de, A. R.; González-Díaz, H.; Molina, R. R.; Uriarte, E. *Proteins, Struct. Funct. Bioinf.* **2004**, *56*, 715.
46. González-Díaz, H.; Bastida, I.; Castañedo, N.; Nasco, O.; Olazabal, E.; Morales, A.; Serrano, H. S.; Ramos de, A. R. *Bull. Math. Biol.* **2004**, *66*, 1285.
47. González-Díaz, H.; Marrero, Y.; Hernández, I.; Bastida, I.; Tenorio, I.; Nasco, O.; Uriarte, E.; Castañedo, N. C.; Cabrera-Pérez, M. A.; Aguila, E.; Marrero, O.; Morales, A.; González, M. P. *Chem. Res. Tox.* **2003**, *16*, 1318.
48. Todeschini, R.; Consonni, V. *Handbook of Molecular Descriptors*; Wiley VCH: Weinheim, Germany, 2000.
49. Kubinyi, H.; Taylor, J.; Ramdsen, C. Quantitative Drug Design. In *Comprehensive Medicinal Chemistry*; Hansch, C., Ed.; Pergamon, 1990; Vol. 4, p 589.
50. Villa, A.; Zangi, R.; Pieffet, G.; Mark, A. E. *J. Comput. Aid. Mol. Des.* **2003**, *17*, 673.
51. *The Theory of Probability*; Gnedenko, B., Ed.; Mir: Moscow, 1978; pp 107–112.
52. Freund, J. A.; Poschel, T. Stochastic Processes in Physics, Chemistry, and Biology. In *Lecture Notes in Physics*; Springer: Berlin, Germany, 2000.
53. Van Waterbeemd, H. Discriminant Analysis for Activity Prediction. In *Method and Principles in Medicinal Chemistry*; R., Manhnhold, Krogsgaard-Larsen, Timmerman, H., Eds.; Chemometric Methods in Molecular Design; Van Waterbeemd, H., Ed.; VCH: Weinhiem, 1995; 265–282.
54. Kowalski, R. B.; Wold, S. Pattern Recognition in Chemistry. In *Handbook of Statistics*; Krishnaiah, P. R., Kanal, L. N., Eds.; North Holland Publishing Company: Amsterdam, 1982; pp 673–697.
55. STATISTICA for Windows release 6.0. Statsoft Inc., 2001.
56. Garcia, A. G.; Horga de la Parte, J. F. Reacciones adversas a los fármacos. In Índice *de especialidades farmacéuticas. Prescripción racional de fármacos*. Médicos S. A., Ed.; INTERCON: Madrid, 1994; pp 155–173.
57. Tilton, C. R.; Ballows, A.; Hohnadel, C. D.; Reiss, F. R. *Mosby-Year Book*; Clinical Laboratory Medicine: USA, 1992; pp 812–994.
58. Loeb, W. F.; Quimby, F. W. *Clinical Chemistry of Laboratory Animals*, 2nd ed.; Taylor & Francis: Philadelphia, PA, 1999.
59. Ping, C.; Hayes, A. Acute Toxicity and Eyes Irritancy. In *Principles and Methods of Toxicology*; Wallace Hayes, A., Ed., 3rd ed.; Raven: New York, 1994.
60. Wiener, H. *J. Am. Chem. Soc.* **1947**, *69*, 17.
61. Estrada, E. *Chem. Phys. Lett.* **2001**, *336*, 248.
62. Marrero-Ponce, Y.; González-Díaz, H.; Romero-Zaldivar, V.; Torrens, F.; Castro, E. A. *Bioorg. Med. Chem.* **2004**, *12*, 5331.
63. Franke, R. *Theoretical Drug Design Methods*; Elsevier: Amsterdam, 1984.
64. Van Waterbeemd, H. Discriminant Analysis for Activity Prediction. In *Method and Principles in Medicinal Chemistry*; Manhnhold, R., Krogsgaard-Larsen, Timmerman, H., Eds.; Chemometric Methods in Molecular Design; Van Waterbeemd, H., Ed.; VCH: Weinhiem, 1995; Vol. 2, pp 265–282.
65. Toropov, A. A.; Toropova, A. P. *J. Mol. Struct. (Theochem.)* **2001**, *538*, 287.
66. Toropov, A. A.; Toropova, A. P. *J. Mol. Struct. (Theochem.)* **2002**, *581*, 11.
67. Toropov, A. A.; Toropova, A. P. *J. Mol. Struct. (Theochem.)* **2003**, *637*, 1.
68. Toropov, A. A.; Schultz, T. W. *J. Chem. Inf. Comput. Sci.* **2003**, *43*, 560.
69. Toropov, A. A.; Toropova, A. P. *J. Mol. Struct. (Theochem.)* **2004**, *676*, 165.
70. Benichou, C.; Celigny, P. S. *Nouv. Rev. Fr. Hematol.* **1991**, *33*, 257.
71. Sasich, D. L.; Sukkari, R. Drug-induced Blood Disorders. In *Applied Therapeutics: The Clinical Use of Drugs*; Koda Kimble, M. A., Lloyd, Y. Y., Kardjan, A. W., Guglielmo, B. J., Eds.; Lippincott Willians & Wilkins: Philadelphia, 2001; Vol. 85, pp 1–21.
72. Cabrera, M. A.; Bermejo, S. *Bioorg. Med. Chem.* **2004**, *22*, 5833.